

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.**

Application Serial Number: 10/518,927A

Source: PCR

Date Processed by STIC: 4/3/06

# ***ENTERED***



PCT

## RAW SEQUENCE LISTING

DATE: 04/03/2006

PATENT APPLICATION: US/10/518,927A

TIME: 08:03:04

Input Set : A:\REVISED\_SEQUENCE\_LISTING.txt

Output Set: N:\CRF4\04032006\J518927A.raw

3 &lt;110&gt; APPLICANT: Merck Patent GmbH

5 &lt;120&gt; TITLE OF INVENTION: DNA SEQUENCE AND PREPARATION OF GRASS POLLEN ALLERGEN Phl p

7 &lt;130&gt; FILE REFERENCE: MERCK-2966

9 &lt;140&gt; CURRENT APPLICATION NUMBER: US 10/518,927A

10 &lt;141&gt; CURRENT FILING DATE: 2004-12-23

12 &lt;150&gt; PRIOR APPLICATION NUMBER: PCT/EP2003/006092

13 &lt;151&gt; PRIOR FILING DATE: 2003-06-11

15 &lt;160&gt; NUMBER OF SEQ ID NOS: 52

17 &lt;170&gt; SOFTWARE: PatentIn version 3.1

19 &lt;210&gt; SEQ ID NO: 1

20 &lt;211&gt; LENGTH: 1503

21 &lt;212&gt; TYPE: DNA

22 &lt;213&gt; ORGANISM: Phleum pratense

24 &lt;220&gt; FEATURE:

W--&gt; 25 &lt;221&gt; NAME/KEY: artificial\_DNA\_sequence

26 &lt;222&gt; LOCATION: (1)..(69)

27 &lt;223&gt; OTHER INFORMATION: DNA sequence derived from sequenced protein

30 &lt;220&gt; FEATURE:

W--&gt; 31 &lt;221&gt; NAME/KEY: native\_DNA\_sequence

32 &lt;222&gt; LOCATION: (70)..(1503)

33 &lt;223&gt; OTHER INFORMATION:

36 &lt;220&gt; FEATURE:

37 &lt;221&gt; NAME/KEY: CDS

38 &lt;222&gt; LOCATION: (1)..(1503)

39 &lt;223&gt; OTHER INFORMATION:

W--&gt; 42 &lt;400&gt; 1

43 tac ttc ccg ccg ccg gct gct aaa gaa gac ttc ctg ggt tgc ctg gtt 48

44 Tyr Phe Pro Pro Pro Ala Ala Lys Glu Asp Phe Leu Gly Cys Leu Val

45 1 5 10 15

47 aaa gaa atc ccg ccg cgt ctg ttg tac gcg aaa tcg tcg ccg gcg tat 96

48 Lys Glu Ile Pro Pro Arg Leu Leu Tyr Ala Lys Ser Ser Pro Ala Tyr

49 20 25 30

51 ccc tca gtc ctg ggg cag acc atc cgg aac tcg cgg tgg tcg tcg ccg 144

52 Pro Ser Val Leu Gly Gln Thr Ile Arg Asn Ser Arg Trp Ser Ser Pro

53 35 40 45

55 gac aac gtg aag ccg atc tac atc gtc acc ccc acc aac gcc tcc cac 192

56 Asp Asn Val Lys Pro Ile Tyr Ile Val Thr Pro Thr Asn Ala Ser His

57 50 55 60

59 atc cag tcc gcc gtg gtg tgc ggc cgc cgg cac ggt gtc cgc atc cgc 240

60 Ile Gln Ser Ala Val Val Cys Gly Arg Arg His Gly Val Arg Ile Arg

61 65 70 75 80

63 gtg cgc agc ggc ggg cac gac tac gag ggc ctc tcg tac cgg tcc ctg 288

64 Val Arg Ser Gly Gly His Asp Tyr Glu Gly Leu Ser Tyr Arg Ser Leu

## RAW SEQUENCE LISTING

DATE: 04/03/2006

PATENT APPLICATION: US/10/518,927A

TIME: 08:03:04

Input Set : A:\REVISED\_SEQUENCE\_LISTING.txt

Output Set: N:\CRF4\04032006\J518927A.raw

65		85		90		95		
67	cag ccc gag gag ttc gcc gtc gtc gac ctt agc aag atg cgg gcc gtg							336
68	Gln Pro Glu Glu Phe Ala Val Val Asp Leu Ser Lys Met Arg Ala Val							
69		100		105		110		
71	tgg gtg gac ggg aag gcc cgc acg gcg tgg gtc gac tcc ggc gcg cag							384
72	Trp Val Asp Gly Lys Ala Arg Thr Ala Trp Val Asp Ser Gly Ala Gln							
73		115		120		125		
75	ctc ggc gag ctc tac tac gcc atc cac aag gcg agt aca gtg ctg gcg							432
76	Leu Gly Glu Leu Tyr Tyr Ala Ile His Lys Ala Ser Thr Val Leu Ala							
77		130		135		140		
79	ttc ccg gcc ggc gtg tgc ccg acc atc ggc gtg ggc ggc aac ttc gcg							480
80	Phe Pro Ala Gly Val Cys Pro Thr Ile Gly Val Gly Gly Asn Phe Ala							
81	145		150		155		160	
83	ggc ggc ggc ttc ggc atg ctg ctg cgc aag tac ggc atc gcg gcc gag							528
84	Gly Gly Gly Phe Gly Met Leu Leu Arg Lys Tyr Gly Ile Ala Ala Glu							
85		165		170		175		
87	aac gtc atc gac gtg aag ctc gtc gac gcc aac ggc acg ctg cac gac							576
88	Asn Val Ile Asp Val Lys Leu Val Asp Ala Asn Gly Thr Leu His Asp							
89		180		185		190		
91	aag aag tcc atg ggc gac gac cat ttc tgg gcc gtc agg ggc ggc ggc							624
92	Lys Lys Ser Met Gly Asp Asp His Phe Trp Ala Val Arg Gly Gly Gly							
93		195		200		205		
95	ggc gag agc ttc ggc atc gtg gtc gcg tgg aag gtg agg ctc ctg ccg							672
96	Gly Glu Ser Phe Gly Ile Val Val Ala Trp Lys Val Arg Leu Leu Pro							
97		210		215		220		
99	gtg ccg ccc acg gtg acc gtg ttc aag atc ccc aag aag gcg agc gag							720
100	Val Pro Pro Thr Val Thr Val Phe Lys Ile Pro Lys Lys Ala Ser Glu							
101	225		230		235		240	
103	ggc gcc gtg gac atc atc aac agg tgg cag gtg gtc gcg ccg cag ctc							768
104	Gly Ala Val Asp Ile Ile Asn Arg Trp Gln Val Val Ala Pro Gln Leu							
105		245		250		255		
107	ccc gac gac ctc atg atc cgc gtc atc gcg cag ggc ccc acg gcc acg							816
108	Pro Asp Asp Leu Met Ile Arg Val Ile Ala Gln Gly Pro Thr Ala Thr							
109		260		265		270		
111	ttc gag gcc atg tac ctg ggc acc tgc caa acc ctg acg ccg atg atg							864
112	Phe Glu Ala Met Tyr Leu Gly Thr Cys Gln Thr Leu Thr Pro Met Met							
113		275		280		285		
115	agc agc aag ttc ccg gag ctc ggc atg aac gcc tcg cac tgc aac gag							912
116	Ser Ser Lys Phe Pro Glu Leu Gly Met Asn Ala Ser His Cys Asn Glu							
117		290		295		300		
119	atg tcg tgg atc cag tcc atc ccc ttc gtc cac ctc ggc cac agg gac							960
120	Met Ser Trp Ile Gln Ser Ile Pro Phe Val His Leu Gly His Arg Asp							
121	305		310		315		320	
123	aac atc gag gac gac ctc ctc aac ccg aac aac acc ttc aag ccc ttc							1008
124	Asn Ile Glu Asp Asp Leu Leu Asn Arg Asn Asn Thr Phe Lys Pro Phe							
125		325		330		335		
127	gcc gaa tac aag tcg gac tac gtc tac gag ccg ttc ccc aag agg gtg							1056
128	Ala Glu Tyr Lys Ser Asp Tyr Val Tyr Glu Pro Phe Pro Lys Arg Val							
129		340		345		350		

## RAW SEQUENCE LISTING

DATE: 04/03/2006

PATENT APPLICATION: US/10/518,927A

TIME: 08:03:04

Input Set : A:\REVISED\_SEQUENCE\_LISTING.txt

Output Set: N:\CRF4\04032006\J518927A.raw

```

131 tgg gag cag atc ttc agc acc tgg ctc ctg aag ccc ggc gcg ggg atc      1104
132 Trp Glu Gln Ile Phe Ser Thr Trp Leu Leu Lys Pro Gly Ala Gly Ile
133          355          360          365
135 atg atc ttc gac ccc tac ggc gcc acc atc agc gcc acc ccg gag tgg      1152
136 Met Ile Phe Asp Pro Tyr Gly Ala Thr Ile Ser Ala Thr Pro Glu Trp
137          370          375          380
139 gcg acg ccg ttc cct cac cgc aag ggc gtc ctc ttc aac atc cag tac      1200
140 Ala Thr Pro Phe Pro His Arg Lys Gly Val Leu Phe Asn Ile Gln Tyr
141 385          390          395          400
143 gtc aac tac tgg ttc gcc ccg gga gcc ggc gcg gcg cca ttg tcg tgg      1248
144 Val Asn Tyr Trp Phe Ala Pro Gly Ala Gly Ala Ala Pro Leu Ser Trp
145          405          410          415
147 agc aag gag atc tac aac tac atg gag cca tac gtg agc aag aac ccc      1296
148 Ser Lys Glu Ile Tyr Asn Tyr Met Glu Pro Tyr Val Ser Lys Asn Pro
149          420          425          430
151 agg cag gcc tac gcc aac tac agg gac atc gac ctc ggg agg aac gag      1344
152 Arg Gln Ala Tyr Ala Asn Tyr Arg Asp Ile Asp Leu Gly Arg Asn Glu
153          435          440          445
155 gtg gtg aac gac gtc tcc acc ttc agc agc ggt ttg gtg tgg ggc cag      1392
156 Val Val Asn Asp Val Ser Thr Phe Ser Ser Gly Leu Val Trp Gly Gln
157          450          455          460
159 aaa tac ttc aag ggc aat ttc cag agg ctc gcc atc acc aag ggc aag      1440
160 Lys Tyr Phe Lys Gly Asn Phe Gln Arg Leu Ala Ile Thr Lys Gly Lys
161 465          470          475          480
163 gtg gat ccc acc gac tac ttc agg aac gag cag agc atc ccg ccg ctc      1488
164 Val Asp Pro Thr Asp Tyr Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu
165          485          490          495
167 atc aaa aag tac tga      1503
168 Ile Lys Lys Tyr
169          500
172 <210> SEQ ID NO: 2
173 <211> LENGTH: 500
174 <212> TYPE: PRT
175 <213> ORGANISM: Phleum pratense
177 <400> SEQUENCE: 2
179 Tyr Phe Pro Pro Pro Ala Ala Lys Glu Asp Phe Leu Gly Cys Leu Val
180 1          5          10          15
183 Lys Glu Ile Pro Pro Arg Leu Leu Tyr Ala Lys Ser Ser Pro Ala Tyr
184          20          25          30
187 Pro Ser Val Leu Gly Gln Thr Ile Arg Asn Ser Arg Trp Ser Ser Pro
188          35          40          45
191 Asp Asn Val Lys Pro Ile Tyr Ile Val Thr Pro Thr Asn Ala Ser His
192          50          55          60
195 Ile Gln Ser Ala Val Val Cys Gly Arg Arg His Gly Val Arg Ile Arg
196 65          70          75          80
199 Val Arg Ser Gly Gly His Asp Tyr Glu Gly Leu Ser Tyr Arg Ser Leu
200          85          90          95
203 Gln Pro Glu Glu Phe Ala Val Val Asp Leu Ser Lys Met Arg Ala Val
204          100          105          110

```

## RAW SEQUENCE LISTING

DATE: 04/03/2006

PATENT APPLICATION: US/10/518,927A

TIME: 08:03:04

Input Set : A:\REVISED\_SEQUENCE\_LISTING.txt

Output Set: N:\CRF4\04032006\J518927A.raw

```

207 Trp Val Asp Gly Lys Ala Arg Thr Ala Trp Val Asp Ser Gly Ala Gln
208      115      120      125
211 Leu Gly Glu Leu Tyr Tyr Ala Ile His Lys Ala Ser Thr Val Leu Ala
212      130      135      140
215 Phe Pro Ala Gly Val Cys Pro Thr Ile Gly Val Gly Gly Asn Phe Ala
216 145      150      155      160
219 Gly Gly Gly Phe Gly Met Leu Leu Arg Lys Tyr Gly Ile Ala Ala Glu
220      165      170      175
223 Asn Val Ile Asp Val Lys Leu Val Asp Ala Asn Gly Thr Leu His Asp
224      180      185      190
227 Lys Lys Ser Met Gly Asp Asp His Phe Trp Ala Val Arg Gly Gly Gly
228      195      200      205
231 Gly Glu Ser Phe Gly Ile Val Val Ala Trp Lys Val Arg Leu Leu Pro
232      210      215      220
235 Val Pro Pro Thr Val Thr Val Phe Lys Ile Pro Lys Lys Ala Ser Glu
236 225      230      235      240
239 Gly Ala Val Asp Ile Ile Asn Arg Trp Gln Val Val Ala Pro Gln Leu
240      245      250      255
243 Pro Asp Asp Leu Met Ile Arg Val Ile Ala Gln Gly Pro Thr Ala Thr
244      260      265      270
247 Phe Glu Ala Met Tyr Leu Gly Thr Cys Gln Thr Leu Thr Pro Met Met
248      275      280      285
251 Ser Ser Lys Phe Pro Glu Leu Gly Met Asn Ala Ser His Cys Asn Glu
252      290      295      300
255 Met Ser Trp Ile Gln Ser Ile Pro Phe Val His Leu Gly His Arg Asp
256 305      310      315      320
259 Asn Ile Glu Asp Asp Leu Leu Asn Arg Asn Asn Thr Phe Lys Pro Phe
260      325      330      335
263 Ala Glu Tyr Lys Ser Asp Tyr Val Tyr Glu Pro Phe Pro Lys Arg Val
264      340      345      350
267 Trp Glu Gln Ile Phe Ser Thr Trp Leu Leu Lys Pro Gly Ala Gly Ile
268      355      360      365
271 Met Ile Phe Asp Pro Tyr Gly Ala Thr Ile Ser Ala Thr Pro Glu Trp
272      370      375      380
275 Ala Thr Pro Phe Pro His Arg Lys Gly Val Leu Phe Asn Ile Gln Tyr
276 385      390      395      400
279 Val Asn Tyr Trp Phe Ala Pro Gly Ala Gly Ala Ala Pro Leu Ser Trp
280      405      410      415
283 Ser Lys Glu Ile Tyr Asn Tyr Met Glu Pro Tyr Val Ser Lys Asn Pro
284      420      425      430
287 Arg Gln Ala Tyr Ala Asn Tyr Arg Asp Ile Asp Leu Gly Arg Asn Glu
288      435      440      445
291 Val Val Asn Asp Val Ser Thr Phe Ser Ser Gly Leu Val Trp Gly Gln
292      450      455      460
295 Lys Tyr Phe Lys Gly Asn Phe Gln Arg Leu Ala Ile Thr Lys Gly Lys
296 465      470      475      480
299 Val Asp Pro Thr Asp Tyr Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu
300      485      490      495
303 Ile Lys Lys Tyr

```

## RAW SEQUENCE LISTING

DATE: 04/03/2006

PATENT APPLICATION: US/10/518,927A

TIME: 08:03:04

Input Set : A:\REVISED\_SEQUENCE\_LISTING.txt

Output Set: N:\CRF4\04032006\J518927A.raw

```

304          500
307 <210> SEQ ID NO: 3
308 <211> LENGTH: 1503
309 <212> TYPE: DNA
310 <213> ORGANISM: Phleum pratense
312 <220> FEATURE:
W--> 313 <221> NAME/KEY: artificial_DNA_sequence
314 <222> LOCATION: (1)..(69)
315 <223> OTHER INFORMATION: DNA sequence derived from sequenced protein
318 <220> FEATURE:
W--> 319 <221> NAME/KEY: native_DNA_sequence
320 <222> LOCATION: (70)..(1503)
321 <223> OTHER INFORMATION:
324 <220> FEATURE:
325 <221> NAME/KEY: CDS
326 <222> LOCATION: (1)..(1503)
327 <223> OTHER INFORMATION:
W--> 330 <400> 3
331 tac ttc ccg ccg ccg gct gct aaa gaa gac ttc ctg ggt tgc ctg gtt      48
332 Tyr Phe Pro Pro Pro Ala Ala Lys Glu Asp Phe Leu Gly Cys Leu Val
333 1          5          10          15
335 aaa gaa atc ccg ccg cgt ctg ttg tac gcg aaa tcg tcg ccg gcg tat      96
336 Lys Glu Ile Pro Pro Arg Leu Leu Tyr Ala Lys Ser Ser Pro Ala Tyr
337          20          25          30
339 ccc tca gtc ctg ggg cag acc atc cgg aac tcg cgg tgg tcg tcg ccg      144
340 Pro Ser Val Leu Gly Gln Thr Ile Arg Asn Ser Arg Trp Ser Ser Pro
341          35          40          45
343 gac aac gtg aag ccg atc tac atc gtc acc ccc acc aac gcc tcc cac      192
344 Asp Asn Val Lys Pro Ile Tyr Ile Val Thr Pro Thr Asn Ala Ser His
345          50          55          60
347 atc cag tcc gcc gtg gtg tgc ggc cgc cgg cac ggt gtc cgc atc cgc      240
348 Ile Gln Ser Ala Val Val Cys Gly Arg Arg His Gly Val Arg Ile Arg
349 65          70          75          80
351 gtg cgc agc ggc ggg cac gac tac gag ggc ctc tcg tac cgg tcc ctg      288
352 Val Arg Ser Gly Gly His Asp Tyr Glu Gly Leu Ser Tyr Arg Ser Leu
353          85          90          95
355 cag ccc gag gag ttc gcc gtc gtc gac ctt agc aag atg cgg gcc gtg      336
356 Gln Pro Glu Glu Phe Ala Val Val Asp Leu Ser Lys Met Arg Ala Val
357          100          105          110
359 tgg gtg gac ggg aag gcc cgc acg gcg tgg gtc gac tcc ggc gcg cag      384
360 Trp Val Asp Gly Lys Ala Arg Thr Ala Trp Val Asp Ser Gly Ala Gln
361          115          120          125
363 ctc ggc gag ctc tac tac gcc atc cac aag gcg agt cca gtg ctg gcg      432
364 Leu Gly Glu Leu Tyr Tyr Ala Ile His Lys Ala Ser Pro Val Leu Ala
365          130          135          140
367 ttc ccg gcc ggc gtg tgc ccg acc atc ggc gtg ggc ggc aac ttc gcg      480
368 Phe Pro Ala Gly Val Cys Pro Thr Ile Gly Val Gly Gly Asn Phe Ala
369 145          150          155          160
371 ggc ggc ggc ttc ggc atg ctg ctg cgc aag tac ggc atc gcg gcc gag      528

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/518,927A

DATE: 04/03/2006  
TIME: 08:03:05

Input Set : A:\REVISED\_SEQUENCE\_LISTING.txt  
Output Set: N:\CRF4\04032006\J518927A.raw

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:7; Xaa Pos. 6  
Seq#:20; Xaa Pos. 8  
Seq#:27; Xaa Pos. 14  
Seq#:28; Xaa Pos. 6  
Seq#:29; Xaa Pos. 3,6,8  
Seq#:30; Xaa Pos. 2,9  
Seq#:33; Xaa Pos. 14  
Seq#:35; Xaa Pos. 14  
Seq#:36; Xaa Pos. 14  
Seq#:37; Xaa Pos. 4,5,14  
Seq#:38; Xaa Pos. 4,5,14  
Seq#:39; Xaa Pos. 4,5,14  
Seq#:40; Xaa Pos. 4,5  
Seq#:41; Xaa Pos. 6  
Seq#:42; Xaa Pos. 3,6,8  
Seq#:43; Xaa Pos. 2,9  
Seq#:46; N Pos. 3,9,15,18,21,24  
Seq#:47; N Pos. 6,12,15,18  
Seq#:48; N Pos. 6,12,18,21

## VERIFICATION SUMMARY

DATE: 04/03/2006

PATENT APPLICATION: US/10/518,927A

TIME: 08:03:05

Input Set : A:\REVISED\_SEQUENCE\_LISTING.txt

Output Set: N:\CRF4\04032006\J518927A.raw

L:25 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:1  
L:31 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:1  
L:42 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:33  
L:42 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:39  
L:313 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:3  
L:319 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:3  
L:330 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:321  
L:330 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:327  
L:601 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:5  
L:607 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:5  
L:618 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:5,Line#:609  
L:618 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:5,Line#:615  
L:896 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7 after pos.:0  
L:1057 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:20 after pos.:0  
L:1164 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:27 after pos.:0  
L:1189 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:0  
L:1210 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:29 after pos.:0  
L:1227 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:30 after pos.:0  
L:1266 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:33 after pos.:0  
L:1294 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:35 after pos.:0  
L:1319 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:36 after pos.:0  
L:1340 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:37 after pos.:0  
L:1357 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:38 after pos.:0  
L:1374 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:39 after pos.:0  
L:1391 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:40 after pos.:0  
L:1408 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:41 after pos.:0  
L:1429 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:42 after pos.:0  
L:1446 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:43 after pos.:0  
L:1484 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:46 after pos.:0  
L:1499 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:47 after pos.:0  
L:1514 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:48 after pos.:0